A Case Study for Educators with ChatGPT and Plato's Allegory of the Cave

Anna Y.Q. HUANG[a], Jain-Wei TZENG[b], Chi-Sheng HUANG[b], Zhi-Qi LIU[c], Bryan Carl TANUJAYA[c] & Owen H.Q. Lu[c*]
[a]*National Central University, Taoyuan*
[b]*National Taichung University of Science and Technology, Taichung*
[c]*National Chengchi University, Taipei*
*owen.lu.academic@gmail.com

Abstract: Generative AI (GenAI) has demonstrated its benefits for student learning outcomes in various educational settings. Due to its diverse applications, it has become a trend for many researchers to explore its usage. In this study, we applied GenAI, specifically ChatGPT, in a workshop designed for 43 university teachers to understand the importance of Plato's Allegory of the Cave in the context of education. The Allegory was chosen because the university teachers came from diverse disciplines, and we needed a subject area that they were not already familiar with. The workshop utilized GenAI in four main aspects: question creation, topic breakdown, problem-solving, and post-class summaries. After the workshop, we used statistical and natural language processing techniques to analyze the teachers' learning outcomes regarding the Allegory of the Cave and the potential of post-class summaries for assessment. The experimental results showed a significant improvement in the teachers' understanding of Plato's Allegory of the Cave. It also confirmed that post-class summaries, as the highest form of knowledge expression, were significantly related to their learning outcomes.

Keywords: ChatGPT, learning outcome, Allegory of the Cave

1. Introduction

Generative Artificial Intelligence (GenAI) has brought a new idea in various scenarios. In healthcare, finance creative arts, or engineering (Cao et al., 2023), the functionality of GenAI to create high-quality content to propose a novel solution is reimagining how we approach problem-solving & creativity. One available functionality on GenAI is that teachers can use the power to produce personalized learning environments that focus on each unique learner, facilitating a more effective approach in the classroom. In this study, we aim to raise attention to applying GenAI to learning activities design and propose a GenAI-collaborative teaching and learning strategy. To this end, we introduced Plato's "Allegory of the Cave," into adult education, specifically the educators and staffs in a university. This allegory depicts people as prisoners in a cave, confusing shadows on the wall for reality and not knowing these are just reflections of real things outside. Such a great metaphor for the limitations of human perception of the importance of education.

In real classrooms, the effectiveness of GenAI must be considered from a more diverse perspective. For example, does GenAI affect students' learning performance once they are overused or misused the functionality of GenAI. Therefore, how to apply GenAI in the classroom correctly becomes the primary motivation of this study. To support and build up an empirical study as a future reference, this study applied GenAI, especially ChatGPT in a workshop. First, there are various GenAI options, for example: Gemini, and Co-polit. However, this study will especially focus on ChatGPT. We select ChatGPT because of its natural language process performance, and easy-to-use features. Therefore, we will survey how to design a learning activity with GhatGPT in the next section. Second, we divided our assessment into two parts: teaching and learning (Chiu, 2023). In the teaching part, we

focused on the student evaluation process, which needs the most effort from teachers especially if they submit text and the context is complex (Garner, 1982). On the other hand, in the learning part, we evaluate students' learning performance when we integrate ChatGPT into students' learning activities.

To summarize, this study will introduce learning activities in a workshop, where students will collaborate with ChatGPT. We will use Plato's Allegory of the Cave as the context of the workshop because most of the participants with less understanding of this concept, subsequently supporting the performance evaluation quality. We divided the evaluation into teaching and learning parts, and the research question is defined respectively as follows:

- **RQ1:** Does GenAI impact workshop participants' learning outcomes in understanding Plato's Allegory of the Cave?
- **RQ2:** How effective are post-class summaries in assessing workshop participants' comprehension of the workshop content?

The subsequent chapters of this paper will be organized as follows: The Literature Review chapter will survey how to apply GenAI in learning activities and methods for assessment. The Research Methods chapter will outline our workshop participants, the design of the learning activities, and how we analyze data. Finally, the Results and Discussion section will introduce the data analysis results and reply to the research question respectively.

2. Literature Review

2.1 State of GenAI in education, lead to the possibility of the followings

GenAI has redefined learning and sparked creativity, facilitating collaborative learning and creative transformation, including personalized learning, improving the quality of educational resources, and optimizing learning and assessment processes (García-Peñalvo, 2024). However, users still need to be mindful of the scope of use and employ relevant tools from a creative perspective to discover new educational methods and enhance human potential (Mishra & Henriksen, 2024). Simultaneously, GenAI can assist teachers in improving the effectiveness of the learning environment. For example, it can generate classroom questions, facilitate interactions, and conduct assessments, effectively saving teachers' time in preparing learning materials and increasing the efficiency of teacher-student interactions (Anderson et al., 2024). However, in evaluating the benefits of GenAI in classroom interactions, scholars have proposed the social construction of technology (SCOT) model, which can observe participants' readiness, gained benefits, and the degree of impact across various learning stages (Oti-Sarpong & Leiringer, 2021; López del Castillo Wilderbeek, 2024). In summary, this paper explores the learning benefits and social construction process of GenAI tools through practical application in university faculty workshops.

2.2 Assessment Methods

This paper employs pre-test and post-test scores, using an independent sample T-test to examine whether the scores achieve statistical significance. During the post-test, participants are required to write a 250-word reflection. This study utilizes Machine Reading Comprehension (MRC) in natural language processing to evaluate machine reading comprehension abilities based on textual reading and question answering (Van Nguyen et al., 2022; Yu & Li 2022). An embedding model is a type of deep learning system that learns low-dimensional feature sets from complex entities (Wang & Lee, 2024). Embeddings are numerical vectors representing word semantics based on their surrounding context, and cosine similarity is calculated using these numerical vectors (Babić et al., 2020). This paper calculates the correlation between post-test scores and written reflections accordingly.

3. Methods

3.1 Participants

The participants in this study consisted of 43 university teachers and staff. Participants' backgrounds are related to science and technology, and the primary subject they taught in information technology, engineering, design, and mathematics. Participants' ages range from 30 to 50 years. To avoid participants' prior knowledge effect our experimental, especially the results of pre-test and post-test, therefore we select the topic of philosophy, specifically Plato's "Allegory of the Cave" as the subject for this workshop.

3.2 Learning activities

To evaluate the effect of GenAI in the classroom, this study conducted a 180-minute workshop with activities as below:
1. **Opening (20 minutes):** explained the workshop background and format. Randomly assign a number card to each participant and use ChatGPT to divide participants into ten groups (prompt: Please randomly divide numbers 1-50 into ten groups).
2. **Pre-test (20 minutes):** deliver a brief background on Plato's Allegory of the Cave, then conduct a pre-test.
3. **Question assignment (10 minutes):** Use ChatGPT to generate ten questions about the Allegory of the Cave (prompt: "Please give me ten open-ended questions about Plato's Allegory of the Cave"). Reordered the questions with the first question asking the first group to introduce a complete story, and the last group to deliver a summary.
4. **Group discussion (40 minutes):** all groups collaborated with ChatGPT to answer the question, then collected results onto a collaborative presentation software.
5. **Group presentation (40 minutes):** all groups presented their answer to the question, and the instructor will lead the discussion to the answers of the pre-test and post-test.
6. **Summarization (30 minutes):** each participant submits a summary of Plato's Allegory of the Cave in Google form, and then the instructor merges all summaries by ChatGPT (prompt: "Here are the summaries provided by the participants of this course. Please consolidate them into a summary of about 200 words explaining the Allegory of the Cave"). Afterward, instructors deliver a final summary.

7. Post-test (20 minutes): Conduct a post-test, then share the pre-test and post-test analysis results.

3.3 Experimental Design

The purpose of the experiment was to measure the participants' understanding of Plato's Allegory of the Cave after the workshop. To this end, the same test with ten questions was conducted before and after the workshop, then we will measure the score difference between the pre-test and post-test. Ideally, the average score will show a significant difference in the post-test, and then we can prove the efficiency of GenAI adoption. This test consisted of fifteen true/false questions about the Allegory of the Cave, generated by using ChatGPT. The workshop instructor then selected ten questions from these, organizing them according to the sequence of events in the story. Moreover, answers and scores won't be shown in the pre-test, but post-test.

3.4 Data preprocess

After the course concluded, we collected each participant's post-class summary to explore the correlation between these summaries and their learning outcomes, we calculate a sim_score by adopting the process as shown in Figure 1. First, we used ChatGPT to consolidate the participants' summaries into a single 250-word summary (prompt: "Please summarize the

following sentences into a 250-word summary"). This consolidated summary included all the key points from the participants' individual summaries. Next, we calculated the similarity between each participant's summary and the consolidated course summary. Due to varying lengths of the participants' summaries, we first padded the sentences to a uniform length. Then, both the padded summaries and the course summary were converted into vectors using an embedding model. We calculated the cosine similarity between these vectors, which we referred to as the summary similarity score. Finally, we performed a correlation analysis between the summary similarity scores and the participants' post-test scores to explore the relationship between the textual summaries and learning outcomes.
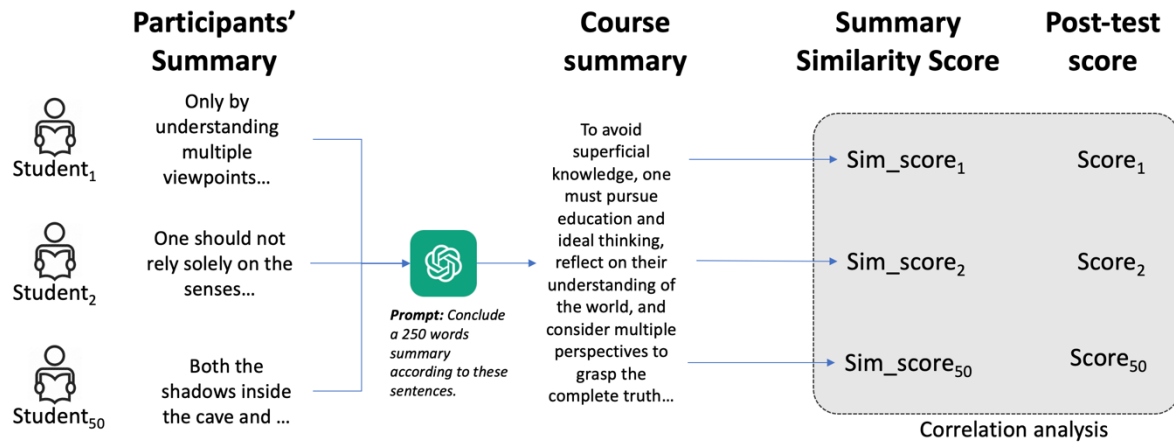

Figure 1. sim_score analysis process

3.5 Evaluation

This study will use quantitative methods to address the two research questions. As for the first research question, we will use a t-test to assess the significance of the differences between the pre-test and post-test scores. For the second question, we will apply Person analysis to evaluate the correlation between the similarity score and students' post-test score.

4. Results and Discussion

4.1 Reply RQ1

To reply RQ1 **(Does GenAI impact workshop participants' learning outcomes in understanding Plato's Allegory of the Cave?)**, this workshop aims to use GenAI to enhance the educational process, and explore the impact of GenAI on learners' learning outcomes for learning new knowledge. Participants were randomly divided into groups and participated in a series of learning activities related to Plato's "Allegory of the Cave" through cooperation with ChatGPT. To measure participants' learning outcomes, this study used pretest and posttest tests (consisting of 10 questions) to measure their understanding of Plato's Allegory of the Cave before and after participating in this workshop activity.

Table1. The result of the independent t-test between pre-test and post-test learning outcomes

Score	Students	Mean/	Standard Deviation	t-value
Pre-test	43	69.22	18.66	-6.96***
Post-test	43	87.91	13.94	

Table 1 indicates that the total number of participants in the workshop's pre-test and post-test was 43 students. In the post-test, the students averaged 69.22 and a standard deviation of 18.66. On the other hand, there was an increase in the post-test score average to

87.91 with a standard deviation of 13.94. The t-value shows the result of -6.92 and p < .001; the negative number from the t-value proves that the scores in the post-test significantly improved from the pre-test. The p-value indicates high significance, signifying a substantial difference in post-test results. It suggests that using GenAI significantly enhanced students' understanding of Plato's "Allegory of the Cave." During the pre-test, the students had limited knowledge of the subject, and utilizing GenAI significantly enhanced their comprehension of the concept. Additionally, GenAI facilitated the student in acquiring knowledge rapidly of a new subject. It also validates that the GenAI was practical and applicable in the learning process through its function to facilitate group discussions, generate questions, and provide summarization.

Given that most participants lack a philosophical background, Plato's "Allegory of the Cave" was chosen as the topic precisely to explore the participants' learning outcome when entering unfamiliar new areas of knowledge. This workshop successfully used Plato's "Allegory of the Cave" to realize a cross-disciplinary learning environment, that is, to explore the impact of GenAI on participants' learning and understanding abilities in a new subject area. This not only helped participants to understanding philosophical content, but also demonstrated how GenAI tools can be used to improve learning in unfamiliar areas.

The Question Assignments and Group presentations at the workshop greatly contributed to the learning experience. This strategy promoted active engagement among students also gain different viewpoints for an in-depth examination of the issue by randomly dividing them into groups and assigning them open-ended questions generated by ChatGPT. The summarization utilizing GenAI assisted the student in consolidating key information and comprehension through accessible language. Overall, through the experiment, the GenAI proved highly effective in facilitating through their ability to engage the students to master new and complex topics quickly. It can be seen from this that GenAI can indeed be used to assist teachers in teaching and students in learning process.

4.2 Reply RQ2

To reply RQ2 **(How effective are post-class summaries in assessing workshop participants' comprehension of the workshop content?)**, the experiment described in Section 3 compared the similarity between individual summaries and the consolidated summary generated by ChatGPT. By analyzing the relationship between the similarity score of individual summaries compared to the consolidated summary and their post-test scores, we aim to determine whether participants who demonstrate a comprehensive understanding of the learning concepts achieve higher scores in the post-test.

Table2. Pearson correlation coefficient results of post-test scores and cosine similarity between individual summary and class summary.

	#	Mean/S.D.	Pearson correlation coefficient	p-value
Post-test score	43	87.91/13.94	0.27	0.09
Cosine Similarity (CS)	43	.32/.009		

#: represent the number of students
S.D.: represent the Standard Deviation

From the results in Table 2, the Pearson correlation coefficient between the post-test scores and the cosine similarity between the individual summary and the class summary is .27, and the p-value is .09, indicating that there is no correlation between the two. In other words, individual summary cannot be used as an accurate measurement tool to understand learning outcome. Results show that summary similarity is not significantly related to traditional test scores (e.g., post-test scores in this study).

Some possible explanations for the above results might be the differing understanding each participant has on the task of writing a summary. Some participants wrote about how

much they knew the allegory and essentially paraphrased the story, while others focused on specific points they learned from the allegory or reflections on the story. While both types of summaries can be generated by participants with a comprehensive understanding of the learning concept who get high score on the post test, the similarity score between an individual summary and the consolidated summary is affected by how closely the participant's summary aligns with the ChatGPT-generated summary. The consolidated summary, which served as a benchmark, included elements from all participants' summaries. Although we have applied padding to reduce the impact of individual summary length on the similarity score, participant's summaries related to understanding the allegory tended to be longer than those focused on personal reflection. This could have caused the consolidated summary to be more resemble to those summaries focused on allegory comprehension when generated by ChatGPT. Further studies could explore different methods for evaluating students' learning outcomes based on text feedback or alternative benchmarks for calculating similarity

5. Conclusion

This study aims to build a reference case for applying GenAI, especially ChatGPT, in the classroom. Four activities out of six were designed in collaboration with ChatGPT. Three of them were applied for grouping, question designing, and course summarization, and the last one was applied under a computer collaborative learning approach. We separated the evaluation of GenAI into learning and teaching parts. The experimental results showed that the participants experienced significant improvement in their learning comprehension after collaborative learning with ChatGPT. On the other hand, ChatGPT has also evaluated the feasibility to be a solution to support grading through participants' post-class summaries. The correlation between summaries and learning outcomes cannot be built by ChatGPT. However, there are potential issues behind this result, and that can be deeply discussed in future studies. For example, we fixed the length of the summary generated by ChatGPT, this made us question what if the summary does not present every detail in the class. More than that, the text-processing models could also be a factor to affect the results, a deep analysis between models to build a robust correlation has to be investigated in our next stage.

Acknowledgements

References

Anderson, O. S., *et al.* (2024). Public health students and instructors weigh in on generative artificial intelligence: Are they on the same page? *Pedagogy in Health Promotion.* https://doi.org/10.1177/23733799241246954

Babić, K., *et al.* (2020). A comparison of approaches for measuring the semantic similarity of short texts based on word embeddings. *Journal of Information and Organizational Sciences, 44*(2), 231–246. https://doi.org/10.31341/jios.44.2.3

Cao, Y., *et al.* (2023). A comprehensive survey of AI-generated content (AIGC): A history of generative AI from GAN to ChatGPT. *arXiv preprint* arXiv:2303.04226. https://doi.org/10.48550/arXiv.2303.04226

Chiu, T. K. (2023). The impact of generative AI (GenAI) on practices, policies, and research direction in education: A case of ChatGPT and Midjourney. *Interactive Learning Environments*, 1–17. https://doi.org/10.1080/10494820.2023.2275687

García-Peñalvo, F. J. (2024). *Generative artificial intelligence and education: An analysis from multiple perspectives.* Ediciones Universidad de Salamanca. https://doi.org/10.14201/e31942

Garner, R. (1982). Efficient text summarization costs and benefits. *The Journal of Educational Research, 75*(5), 275–279. https://doi.org/10.1080/00220671.1982.10885269

López del Castillo Wilderbeek, F. L. (2024). Generative artificial intelligence: Technological determinism or socially constructed artifact. *Palabra Clave, 27*(1). https://doi.org/10.5294/pacla.2024.27.1.7

Mishra, P., & Henriksen, D. (2024). Creative dialogue with generative AI: Exploring the possible with Ron Beghetto. *TechTrends, 68*(3), 395–401. https://doi.org/10.1007/s11528-023-00894-0

Oti-Sarpong, K., & Leiringer, R. (2021). International technology transfer through projects: A social construction of technology perspective. *International Journal of Project Management, 39*(8), 902–914. https://doi.org/10.1016/j.ijproman.2021.06.003

Van Nguyen, K., *et al.* (2022). New Vietnamese corpus for machine reading comprehension of health news articles. *ACM Transactions on Asian and Low-Resource Language Information Processing, 21*(5), 1–28. https://doi.org/10.1145/3494807

Wang, Q., & Lee, K. A. (2024). Cosine scoring with uncertainty for neural speaker embedding. *IEEE Signal Processing Letters.* https://doi.org/10.1109/LSP.2024.3258564

Yu, C., & Li, X. (2022). SSAG-Net: Syntactic and semantic attention-guided machine reading comprehension. *Intelligent Automation & Soft Computing, 34*(3), 1897–1910. https://doi.org/10.32604/iasc.2022.021399